Self-supervised Image Enhancement Network: Training with Low Light Images Only

Yu Zhang, Xiaoguang Di, Bin Zhang, and Chunhui Wang

Abstract—This paper proposes a self-supervised low light image enhancement method based on deep learning. Inspired by information entropy theory and Retinex model, we proposed a maximum entropy based Retinex model. With this model, a very simple network can separate the illumination and reflectance, and the network can be trained with low light images only. We introduce a constraint that the maximum channel of the reflectance conforms to the maximum channel of the low light image and its entropy should be largest in our model to achieve self-supervised learning. Our model is very simple and does not rely on any well-designed data set (even one low light image can complete the training). The network only needs minute-level training to achieve image enhancement. It can be proved through experiments that the proposed method has reached the state-of-the-art in terms of processing speed and effect.

Index Terms—Low Light Image Enhancement, Self-supervised Learning, Max Entropy, Retinex.

1 Introduction

Recently, various algorithms based on deep learning have achieved surprising results in some image processing and computer vision tasks, such as object detection [1], [2], [3], [4], image segmentation [4], [5], [6], etc. One important reason for the rapid development of deep learning in these tasks is that we can obtain a large number of data sets with clear and unambiguous labels. In these tasks, although the construction of the data set requires some cost, it is still acceptable, also on the Internet, a large number of open source data sets can be found for these tasks to support the training of the network. However, in low-level image processing tasks such as low light image enhancement, image dehazing, and image restoration, etc., it is difficult to obtain a large number of true input/label image pairs.

As for low light image enhancement task, in the previous work, some solutions such as synthesizing low light images [7], using different exposure time images to obtain data [8], and so on, have achieved good visual effects. However, there are still two problems with those methods. One is how to ensure that the pre-trained network can be used for images collected from different devices, different scenes, and different lighting conditions rather than building new training data set. The other is how to determine whether the normal light image used for supervision is the best, there can be lots of normal light images for a low light image. Usually, the builder of the data set gets the normal light images by experience or artificial adjustment, which will cost lots of time and energy and we cannot make sure that the enhanced image can show the information contained in the low light image to the greatest extent with those normal light images.

For those two questions, this paper proposes a self-supervised low light image enhancement network based on information entropy theory and Retinex model, and achieves the state-of-the-art in terms of enhancement quality and efficiency. In this paper, the only data we need are the low light images, without any paired or unpaired normal light images. To our knowledge, this is the first fully self-supervised image enhancement method based on deep learning. The proposed method does not rely on a well-designed complex network structure, only with a simple fully convolutional neural network (CNN) as shown in Fig.2 and minute-level training, we can complete low-light image enhancement tasks.

There are some image enhancement networks based on Retinex model [9], [10], but they all require paired data, and then use the assumptions that images captured in different light conditions should have the same reflectance and the illumination map should be smooth to decompose low light images into corresponding reflectance and illumination map. Similar to these works, we also use a network to decompose low light image into reflectance and illumination, but unlike those previous works, we use self-supervised methods to train the network. Only low light images are required (even a single low light image) for training, then we can get the reflectance with good visual effects, and it can be treated as an enhanced image.

We think that low light image enhancement task is to display the information contained in low light images in a more intuitive way, rather than creating new information. At the same time, according to the entropy theory, images whose histogram are uniform distribution have the maximum entropy and contain the most information. Based on the above analysis, we propose an assumption that the histogram distribution of the maximum channel of the enhanced image should conform to the histogram distribution of the maximum channel of the low light image after his-

- *Y. Zhang, B. Zhang and C. Wang are with the Department of Electronic Science and technology, Harbin Institute of Technology, Harbin 150001,China.(E-mail: hitzhangyu@qq.com;teamup@yeah.net;wang2352@hit.edu.cn)*
- *X. Di is with the Department of Control Science and Engineering,Harbin Institute of Technology, Harbin 150001,China.(E-mail: dixiaoguang@hit.edu.cn)*

togram equalization. With this assumption, the loss function can be designed without normal light images, and it can not only retain the authenticity of the enhanced image, but also ensure that the enhanced image has sufficient information. The proposed method does not have any dependence on the way of acquiring low light images, and the training process is completely self-supervised, so the method proposed in this paper has good generalization ability, even if the pre-trained network is not well enough in new environment, retraining or fine-tuning without building paired/unpaired normal light images data set is possible for the network. Our contributions include:

- We propose a new maximum entropy based Retinex model, and give its theoretical source.
- Combined with deep learning, we propose a self-supervised low light image enhancement network, which can complete the training with even one single low light image.
- The proposed method only requires minute-level training and has a good real-time performance. We verify the enhancement effect and stability of the algorithm through some experiments and objective indexes.

2 Related Works

Our method mainly comes from histogram equalization, model-based methods, and deep learning based image enhancement methods.

Histogram Equalization

In low light image enhancement tasks, Histogram Equalization(HE) is the most simply and wildly used method. It can let the histogram of the enhancement image have a uniform distribution to get the maximum entropy. However, HE cannot avoid the problems of details disappearance (over enhancement, under enhancement), poor color restoration, noise amplification and so on.

To solve those problems, various improved algorithms are proposed, such as Adaptive Histogram Equalization(AHE) [11] and Contrast-limited Adaptive Histogram Equalization(CLAHE) [12] for details, Hue-preserving color image enhancement [13] for hue preserving, Brightness Bi-Histogram Equalization Method (BBHE) [14], Dualistic Sub-Image Histogram Equalization Method (DSIHE) [15] for brightness preserving, etc. In [16] and [17], the method considering the relationship between adjacent pixels and large gray-level difference is proposed. Although many improved methods have been proposed, there are still many problems in applying histogram equalization directly to image enhancement.

Model Based Image Enhancement Method

Among the model-based low-light image enhancement methods, there are mainly based on the dehazing model [18] and Retinex model [19]. The method based on the dehazing model is mainly based on the discovery that the low light image is similar to the haze image after inversion. Dong et al. proposed an enhancement method that performs the dehazing operation after inverting the low light image and then inverts the image back [18]. Some studies have extended these works [20], although these methods have achieved some good effect, they lack corresponding physical model, which limits the application of the method in various scenes.

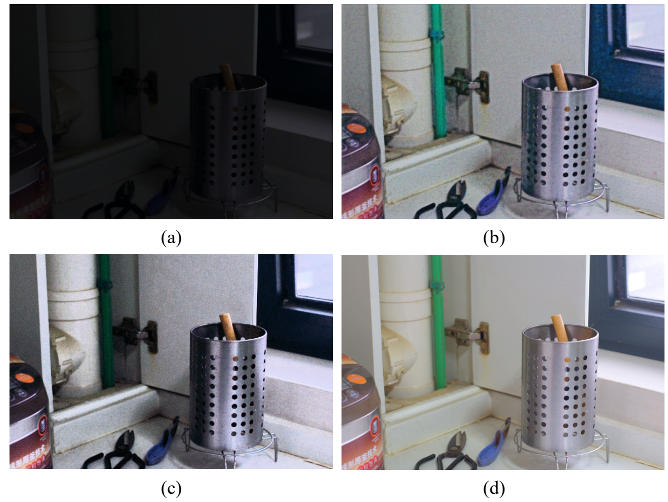

Fig. 1. The enhancement results by proposed method. (a) Input. (b) The network was trained 200 epochs with low-light images without (a). (c) The network was trained 10000 times with image (a) only. (d) Reference

According to Retinex theory, the collected images can be decomposed into illumination and reflectance, but this is a highly ill-posed problem to obtain them from low light images only. Therefore, other constraints must be introduced: the early researches like single-scale Retinex [21] model and the multi-scale Retinex [22] model, only use the constraint that the illumination map is smooth to solve the problem. The captured image is smoothed by using Gaussian filters with one or more scales to obtain illumination, however, the enhanced image often have unreal phenomena such as over-enhancement and whitening. [23] proposes a Bright-Pass Filter that preserves natural characteristics. [24] performs global illumination adjustment and local contrast enhancement on the initially estimated illumination map. Although they have obtained some good effects, without considering the structural characteristics, information loss is prone to occur in areas with rich details. LIME [25] introduces a filter considering the structure characteristic to smooth the illumination map and uses BM3D to denoise the enhanced image. Although those methods are proposed to maintain image details and naturalness, neither before nor after enhancement tasks, the denoising process will still cause blur or loss of details.

It is difficult to add some additional priors to those methods based on Retinex model only, which leads to the problems of noise, halo, detail preservation and so on, so in recent years, many algorithms based on the variational Retinex model are proposed. Kimmel et al. [26] first proposes a variational Retinex model, and uses L_2 regularization to obtain a smooth illumination map. Fu et al. [27] introduces the bright channel prior to the variational Retinex model to suppress the halo effect. Park et al. [28] proposes a weighted L_2 regularization to constraint reflectance image, which has a slight noise suppression effect. Fu et al. [29] proposes a L_2-L_p norm to constrain the illumination map and keep more details. Although these variation based methods have achieved good results, it is

very time-consuming to process images due to the need of multiple iterations to solve the variational equation. Even with Fast Fourier Transform(FFT), it is difficult to ensure the real-time.

In addition, in those model-based low light image enhancement algorithms, spatial smoothing prior [30] and its improvement are mostly used to constrain the illumination map, and there is no constraint on the contrast information of the reflectance. In this paper, we use the maximum information entropy to constrain the reflectance image, so as to further improve its contrast information.

Learning based methods

Learning based methods have achieved good results in some low-level image processing tasks, such as image denoising [31], [32], super-resolution reconstruction [33], [34],restoration [35], [36], etc. However, most of the current algorithms based on deep learning are supervised, and it is difficult to obtain both degraded and normal images in those low-level image processing tasks. It is proposed to synthesize low light image data with normal light image for training in some researches. For example, LLNET [37] is the first work to use deep learning to solve image enhancement problem, it proposes to train the networks with synthetically noisy and dark images separately, but it does not consider the natural images characteristic. In [38] and [39], gamma transformation is applied to natural image patches to generate low light image patches for training, but they do not consider other degradation of the real collected low light images like noise, color changing, etc. In MSR-net [40], high quality (HQ) data are obtained by artificial selection and Photoshop, and low light images are obtained by processing HQ data with random brightness and contrast reduction and gamma transformation. The data obtained by those methods seems to look like low light images, however, it is difficult to truly reflect the characteristics of low light images, such as noise, overexposed and underexposed areas existed in the same image, etc.

In order to solve this problem ,some methods propose to use real low light images for training. In [41], a large multi-exposure image database is established, and the reference images are obtained by combing different exposure images and subjective selection. Retinex-net [9] tries to obtain the low/normal light image pairs through adjusting the exposure time, and achieves good enhancement effects, but the exposure time is still artificially determined, and it is difficult to choose the best exposure time to get a reference image. In [8], it introduces a parameter to link two images with different exposure time, and with end-to-end training, it can well deal with the noise problem, but it can only be used for raw images. In [42], a light adjustment network is introduced to link paired images with any different exposure time, which solves the problem in acquisition of normal light images. However, in practical applications, if we want to get better images, we may need to choose a hyper-parameter for each low light image.

Although these deep learning based methods have achieved good visual effects in low light image enhancement, they are all based on paired images, and the cost of building training data is so high, and they do not solve the two problems we mentioned before, i.e. how to obtain an optimal reference image and how to ensure the adaptability of the method to new environments or new equipments.

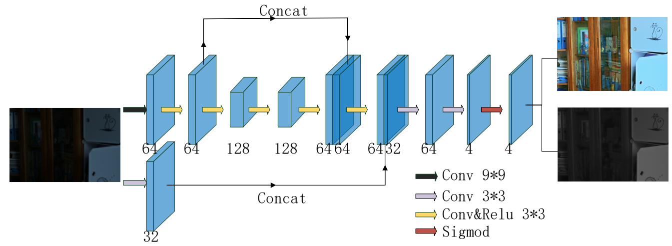

Fig. 2. Structure of the Self-supervised image enhancement network

3 Method

3.1 Maximum Entropy Based Retinex model

Base on Retinex model, an image can be decomposed into reflectance and illumination map as follows:

$$S = R \circ I \tag{1}$$

where S represents the captured image, R represents the reflectance and I represents the illumination map. This is a highly ill-posed problem, its solution needs additional prior. According to Bayesian formula, the problem can be expressed as follows:

$$p(R, I \mid S) \propto p(S \mid R, I)p(R)p(I) \tag{2}$$

Where, $p(R, I \mid S)$ is posterior probability, $p(S \mid R, I)$ is the class conditional probability, and $p(R)$ and $p(I)$ are prior probabilities of reflectance and illumination. Existing methods generally add the prior probabilities $p(R)$ and $p(I)$ to find the maximum posterior probability, and estimate the reflectance and illumination.

By calculating the negative logarithm of equation (2), the problem of image enhancement can be transformed into the form of three distance terms, as can be seen in formula (3):

$$\underset{R,S}{min} l_{rcon} + \lambda_1 l_R + \lambda_2 l_I \tag{3}$$

Where, l_{rcon} represents reconstruction loss, l_R represents reflectance loss, and l_I represents illumination loss. λ_1 and λ_2 are weight parameters.

In this paper, we use the L_1 norm to constrain all the losses, we do not compare the impact of L_1, L_2, SSIM and other loss functions on low level image processing tasks, there are some related studies such as [43]. The reconstruction loss l_{rcon} can be expressed as:

$$l_{rcon} = \|S - R \circ I\|_1 \tag{4}$$

As for the reflectance loss, different from the existing methods only using $\|\triangle R\|_1$ [28], [44], we propose a new distance measurement method for reflectance loss based on the following reasons:

- For image enhancement task, the processed image should have enough information
- The processed image shall conform to the original image information

- Histogram equalization can greatly improve the information entropy of image

Based on the above considerations, we propose equation (5) as the loss of reflectance image, which also uses L_1 loss:

$$l_R = \left\| \max_{c \in R,G,B} R^c - F(\max_{c \in R,G,B} S^c) \right\|_1 + \lambda \left\| \triangle R \right\|_1 \quad (5)$$

Where, $F(X)$ means the histogram equalization operator to image X. λ is weight parameters. This loss function means that maximum channel of the reflectance should conform to the maximum channel of the low light image and has the maximum entropy. There are three main reasons why we choose the maximum channel to constrain. Firstly, for a low light image, the maximum channel has the greatest impact on its visual effect. Secondly, if other channels are selected, there is no doubt that saturation will occur according to the prior that the maximum channel must be greater than the other two channels. Thirdly, if we choose one of the color channel, such as R, G or B channel, it is obviously not in line with the natural image.

For the illumination loss, we adopt the structure-aware smoothness loss proposed in [9]:

$$l_I = \left\| \triangle I \circ exp\left(-\lambda_3 \triangle R\right) \right\|_1 \quad (6)$$

It is proposed that equation (6) can make the illumination loss aware of the image structure in [9]. And the loss means that the original TV function $\left\| \triangle I \right\|_1$ is weighted with the gradient of reflectance.

From the equation (3) to equation (6), we get the maximum entropy based Retinex model, as can be seen in equation (7):

$$\begin{aligned} Z = &\left\| S - R \circ I \right\|_1 + \lambda_1 \left\| \max_{c \in R,G,B} R^c - F(\max_{c \in R,G,B} S^c) \right\|_1 \\ &+ \lambda_2 \left\| \Delta I \circ \lambda exp\left(-\lambda_3 \triangle R\right) \right\|_1 \\ &+ \lambda_4 \left\| \triangle R \right\|_1 \end{aligned} \quad (7)$$

Variational methods or FFT are generally used to solve equation (7) with L_2 loss, however, they both need multiple iterations which will bring time consumption problems, and with more constraints, the solution will be more complicated. In order to enhance the image in real time, we propose a solution based on deep learning. The network uses equation (7) as the loss function. We can find that in equation (7), there is only low light images, so the network can be trained through a self-supervised way.

The values of λ_1,λ_2,λ_3,λ_4 are 0.1, 0.1, 10 and 0.01 in this paper. The influence of values of λ_1 and λ_2 is not so obvious in visual effect that we just choose 0.1, the value of λ_3 comes from [9]. As for λ_4, in our experiments, we found that it can be used to control the noise. When its value increases, the noise decreases, and at the same time, the image will be more blurry. Through some experiments, we choose 0.01 for λ_4 and if $\lambda_4 = 0.1$, the enhanced image will appear obvious blur.

TABLE 1
Self-supervised image enhancement network

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
RGB&maxR,G,B	Conv&ReLU	3×3	32	1	Conv0
RGB&maxR,G,B	Conv	9×9	64	1	Conv
Conv	Conv&ReLU	3×3	64	1	Conv1
Conv1	Conv&ReLU	3×3	128	2↓	Conv2
Conv2	Conv&ReLU	3×3	128	1	Conv3
Conv3	Conv&ReLU	3×3	64	2↑	Conv4
Conv4&Conv1	Concat	-	128	-	Conv5
Conv5	Conv&ReLU	3×3	64	1	Conv6
Conv6&Conv0	Concat	-	96	-	Conv7
Conv7	Conv	3×3	64	1	Conv8
Conv8	Conv&ReLU	3×3	4	1	Conv9
Conv9	Sigmoid	-	4	-	R&I

3.2 Self-supervised Network Based Solution

If we use the variational methods or FFT to solve the model proposed from equation (3) to equation (7), then it means that we need to carry out the same iterative processing for each low light image, which will not only bring time-consuming problems, but also the iteration times for each low light image may be uncertain, which is almost a disaster in many real applications. At the same time, this kind of solution can not take advantage of big data, the previous data processing can do nothing helpful to the new data processing.

In the previous deep learning based researches, due to the lack of models that can support self-supervised training, only the paired or unpaired low/normal light images collected in advance can be used to complete the network training. However, the data collected in advance can not contain all the real low light situations, such as different environments, devices, or degradation problems, etc., which also limits the application scope of the pre-trained network. After all, it is impossible to build the data set when we are using them.

However, based on the model proposed from equation (3) to equation (7), we can achieve the self-supervised training, which means that we can build the data set online and avoid the problem of applicability. And compared with the supervised learning whose supervisor is selected by artificial method, the model based on maximum entropy can ensure that the enhanced image has enough information entropy.

We only need a very simple CNN structure to achieve the decomposition of the illumination and reflectance. The specific structure of the CNN we finally adopted is shown in Fig.2. The input of the network is low light image and its maximum channel, after some convolution and concat layers, reflectance and illumination can be gotten with a sigmod layer. Table 1 is the specific information of each layer of the network.

In fact, we have experimented with different network structures, and the stacking of convolutional layers and a sigmod layer can also produce acceptable results. However, if we add some concat layers, the enhancement results will become clearer. It can be seen that we use down-sampling and up-sampling in the network, its prime function is to reduce the noise. In some experiments, we find that adding the down-sampling layer will make the image blur, however, it will reduce the noise too.

4 Experiment

We use the LOL database [9] which contains 500 low/normal light image pairs, 485 of which are used for training and images size are $400 * 600$. Note that during the training process, we only use natural low light images and do not use synthetic data and normal light images. During the training process, our batch size is set to 16 and the patch size is set to 48 * 48. We use Adam stochastic optimization [45] to train the network and the update rate is set to 0.001. The training and testing of the network are completed on a Nvidia GTX 2080Ti GPU and Inter Core i9-9900K CPU, and the code is based on the tensorflow framework.

In section 4.1, we introduce some objective evaluation indexes. In section 4.2, we measure the influence of the training times on loss and evaluation indexes. In section 4.3, we measure the stability of the algorithm through repeated experiments. In section 4.4, we compare our algorithm with some existing methods. In section 4.5, we give some enhancement results when the network is trained with one single low light image.

4.1 Evaluation Indexes

There are many indexes with or without reference that can be used to evaluate the quality of the enhanced image. However, the constrain we use in this paper does not conform to the natural image characteristics, so it is difficult for us to evaluate the enhanced image accurately with those existing evaluation indexes. In this paper, we use gray entropy (GE), color entropy (CE, color entropy is the sum of entropy of R,G,B channels), gray mean illumination (GMI), gray mean gradient (GMG), LOE [23], NIQE [46], PSNR, SSIM to evaluate the enhanced image. It should be noted that these indexes can only reflect the image quality in some aspects, which are not completely consistent with the evaluation results given by the human visual system. The LOE$_{low}$ and LOE$_{high}$ are calculated with low and high light images respectively.

4.2 The Influence of Training Times

We use 485 low light images in the LOL dataset for training, and 15 for testing. Considering that our method is self-supervised, it lacks an absolute reference, and some parameters and constraints in our loss function come from the individual experience. We cannot determine whether our training has reached the best through the change of the loss. So we train the network for 1,000 epochs, and process the testing data every 20 training epochs and use those indexes to evaluate the training results of the network.

Fig.3 and Fig.4 show the change of loss and indexes with the increase of training times. It can be seen that the loss falling fast at the beginning. On our GPU, it takes less than 0.65s to train one epoch. Fig.5 shows enhancement results of low light images in the testing data with different training times. We only selected the results of the first 200 epochs to display. It can be seen that as the training goes, some indexes which can reflect the image clarity such as entropy and gradient increase, however, the gap between the enhanced images and reference images is also growing. That is caused by noise, although the image becomes more and more clear as the training goes, at the same time, the noise keep increasing too. In order to keep balance between clarity and noise, we just stop training after 200 epochs. And in our experiment, if the training epochs keep increasing more than about 1000 epochs, there will be artifacts in some testing images like [8]. Early stopping is a reasonable method to avoid the noise and artifacts.

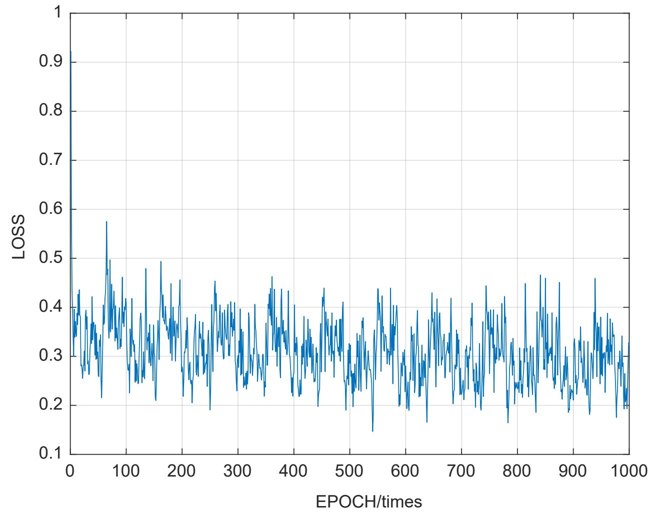

Fig. 3. The training loss with 1000 epochs

4.3 Repeated Learning Stability

Due to the characteristics of learning based methods, in most of cases, we cannot reproduce the optimal results. So we repeat the experiments many times to evaluate the repeatability of the method. In every experiment, we train the network with 200 epochs, and evaluate the network on testing data through indexes mentioned in section 4.1. Fig.6 and Fig.7 show the evaluation indexes and some enhancement results in different experiments respectively. It can be seen that there are large fluctuations in some indexes, like LOE、GMG and NIQE, however, the changings of enhancement results are not so obvious in most experiments. In the fifth experiment, the color of enhanced images are lighter than others. We think that the differences of enhancement results may come from the L_1 loss functions and the difference among the training data in every experiment.

In every experiment, the only difference is training patch, which seems to have an impact on training results. Those training patches are randomly selected and cropped, and considering the training times and the large size difference between images and patches, they are only a small part of training images. At the same time, we use the L_1 loss for training, compared with L_2 loss, L_1 loss may have multiple solutions, and its solutions will be highly affected by training data. When training data changes, the results may also change greatly. However, from visual effects, the method proposed in this paper is relatively stable.

4.4 Comparisons with Existing Algorithms

We have also compared our algorithm with some existing classic and state-of-the-art methods, including HE, MSR

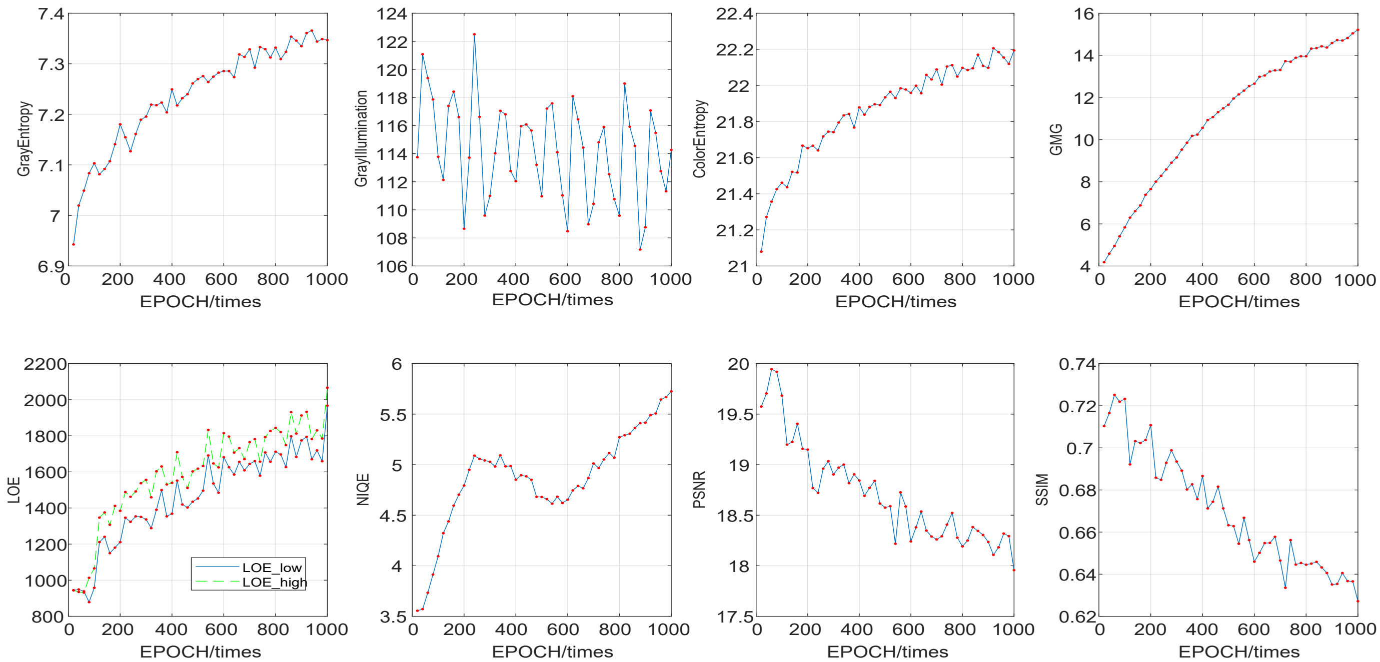

Fig. 4. Evaluation indexes on the testing data by different training epochs

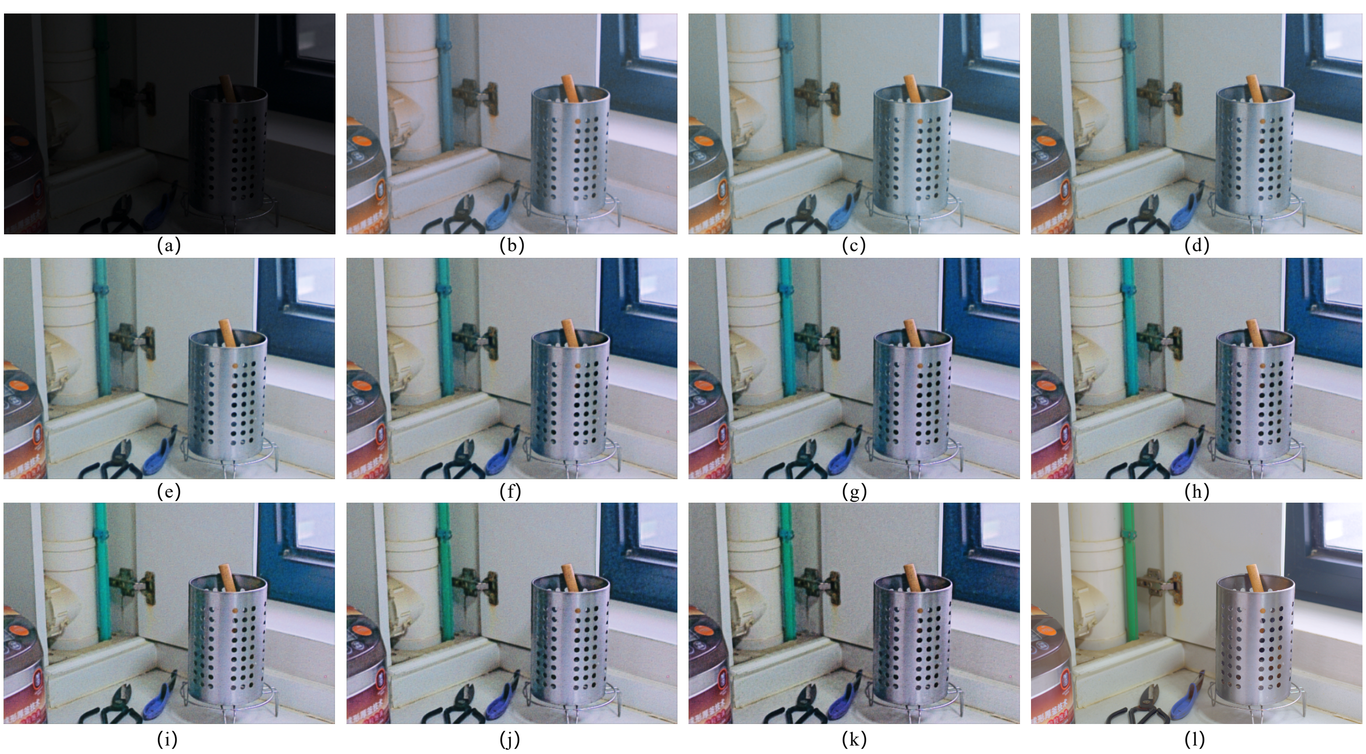

Fig. 5. The enhancement results by different training epochs (a) Original. (b) Epoch = 20 (c) Epoch = 40. (d) Epoch = 60. (e) Epoch =80. (f) Epoch =100. (g) Epoch =120. (h) Epoch =140. (i) Epoch =160. (j) Epoch =180. (k) Epoch =200. (l) Reference.

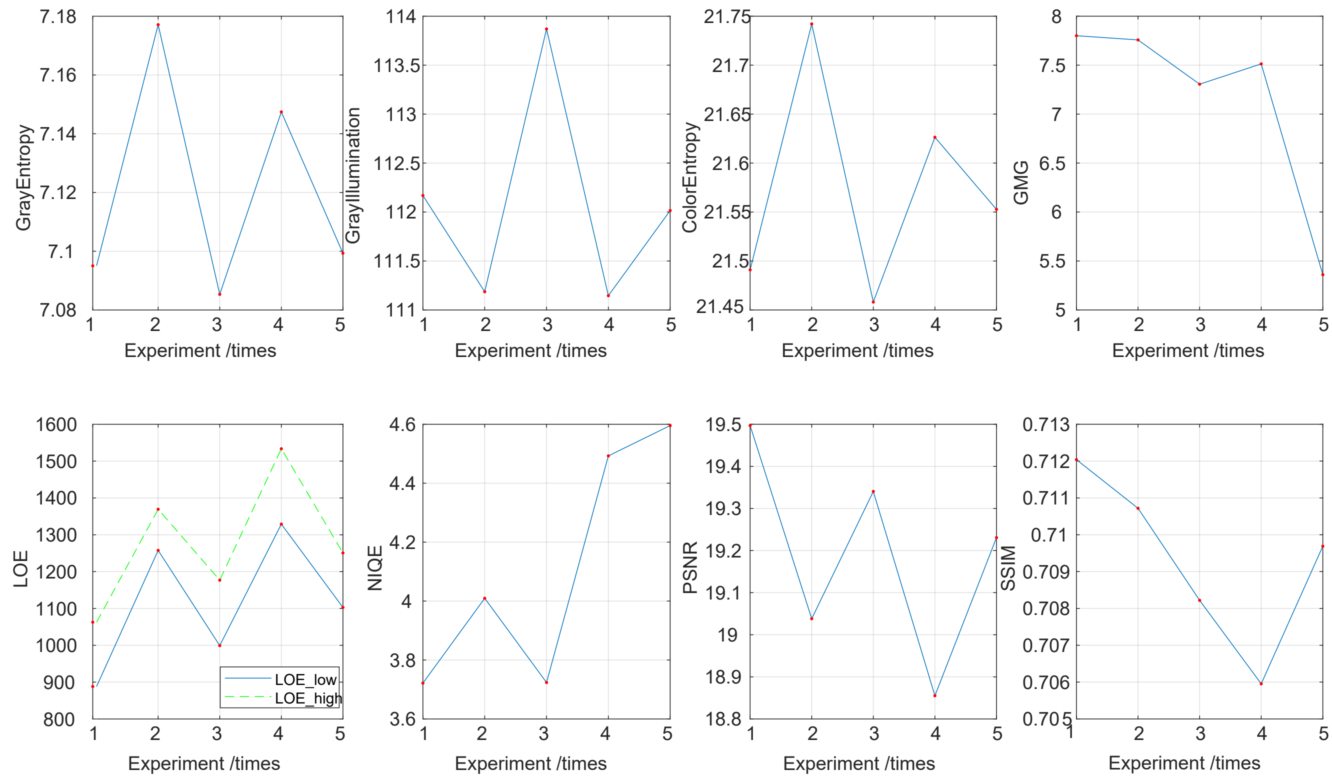

Fig. 6. Index changes in repeated experiments, 200 epochs training in each experiments

Metrics	GE	GMI	CE	GMG	LOE$_{low}$	LOE$_{high}$	NIQE	PSNR	SSIM	Time
HE	**7.785**	105.3	**23.21**	17.12	**505.3**	**898.6**	5.201	15.81	0.5607	0.0158
MSR	6.947	134.7	17.61	16.93	540.3	950.1	8.008	16.69	0.5262	0.0911
NPE	7.070	96.67	20.787	20.38	1607.7	1867.8	9.135	16.97	0.5894	4.71
MF	6.937	85.99	20.16	17.47	840.9	1197.7	9.713	16.97	0.6049	0.128
SRIE	6.296	50.33	18.62	9.380	952.0	1291.4	7.535	11.86	0.4978	3.45
LIME	7.564	114.9	21.72	**23.46**	1303.5	1543.7	9.127	16.76	0.5644	0.211
Gladnet	7.116	119.0	21.52	9.918	902.5	1205.5	6.797	**19.72**	0.7035	0.0212
Retinex-Net	6.835	110.2	21.13	24.00	1990.1	1988.8	9.730	16.77	0.5594	0.0207
$L_2 - L_p$	6.419	51.92	19.16	8.704	933.4	1250.4	6.234	12.15	0.5103	5.58
Ours	7.180	108.7	21.65	7.653	1210.8	1384.1	4.793	19.15	**0.7108**	**0.0145**
Ours-single	7.030	88.08	20.94	10.27	529.0	1028.2	**4.422**	14.18	0.5169	**0.0145**
Reference	7.040	115.5	21.31	6.910	921.9	-	4.253	-	1	-

TABLE 2
Quantitative comparison on LOL dataset in terms of PSNR, SSIM, LOE$_{low}$, LOE$_{high}$, and NIQE. The best results are highlighted in bold.

[22], LIME [25], MF [24], NPE [23],SRIE [47], Gladnet [48], Retinex-Net [9], L_2-L_p [29].

The 15 images from the LOL dataset are used for testing to get the objective indexes, and Fig.8 and Fig.9 show some enhancement results by different methods, and Table 2 displays the objective indexes and time consumption of those methods. The low light image of Fig.10 is from the LIME [25] dataset. All the results of our method come from a randomly selected experiment.

SSIM is generally used to measure the structural similarity between two images. NIQE is a non reference image quality evaluation method. These two indexes can show that our method has good structural similarity and image quality after processing.

In CE, GE and GMG, we can see that our method is lower than some methods. Although the larger these indexes are, the more abundant information these images have and the clearer these images are, we also need to consider that these indexes will be highly affected by noise. It can be seen that compared with most of the methods, our method is closer to the reference image in these indexes.

In LOE$_{low}$ and LOE$_{high}$, our method does not perform well. This is probably because that the overexposure area and underexposure area may exist in low light images or the reference images at the same time, and the model proposed in this paper can avoid this problem to a certain extent, which leads to our poor performance in these two indexes. As shown in the lower left corner of Fig.7, there are overexposure areas in the reference image. If we use such reference image for training, we can not promise that the training results will not be over exposed. It can also explain that it is difficult to obtain the optimal reference image by

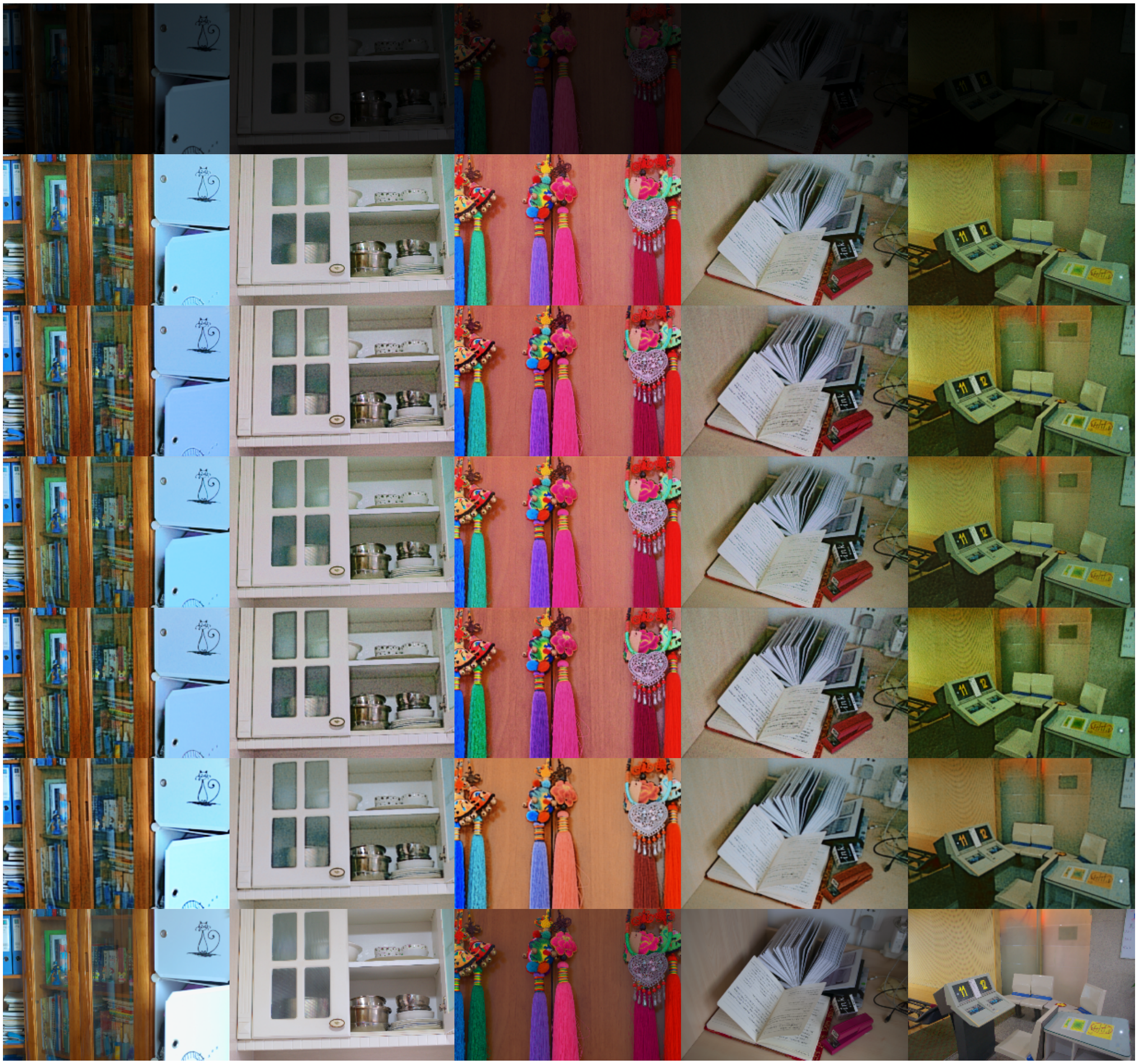

Fig. 7. The first row is the original low light image. The 2-6 rows are different experiments. The last row is the reference normal light image

adjusting the exposure time.

In PSNR, although our method is lower than Gladnet [48], we need to pay attention to that our method does not consider the reference image in training, so it is difficult to ensure the similarity of the enhanced image and the reference image in brightness, which has a certain impact on PSNR.

Although our method can not achieve the best results in all indexes, in terms of visual effect and some important indexes, our method achieves the state-of-the-art. Compared with HE method, our method has a slight denoising effect and can better maintain the structure information and color information. As shown in Table 2, the PSNR and SSIM of our method is higher than the HE method. HE method is not suitable for heavy noise environment and it can be seen in Fig.8 and Fig.9, there are still dark areas in the image after directly using HE method, which is caused by the theory of histogram equalization itself. Compared with the model-based methods, our method cost less running time. It can be seen in Table 2, our method is more than 6x faster than MSR [22] which has the least time among model-based method. Compared with the learning based methods, our method does not need to build data sets carefully, which can save a lot of time and energy and has better applicability for new environment and equipment.

4.5 Training with Single Low Light Image

At the same time, in order to further evaluate the performance of the method proposed in this paper, we do an experiment that train the network with single low light

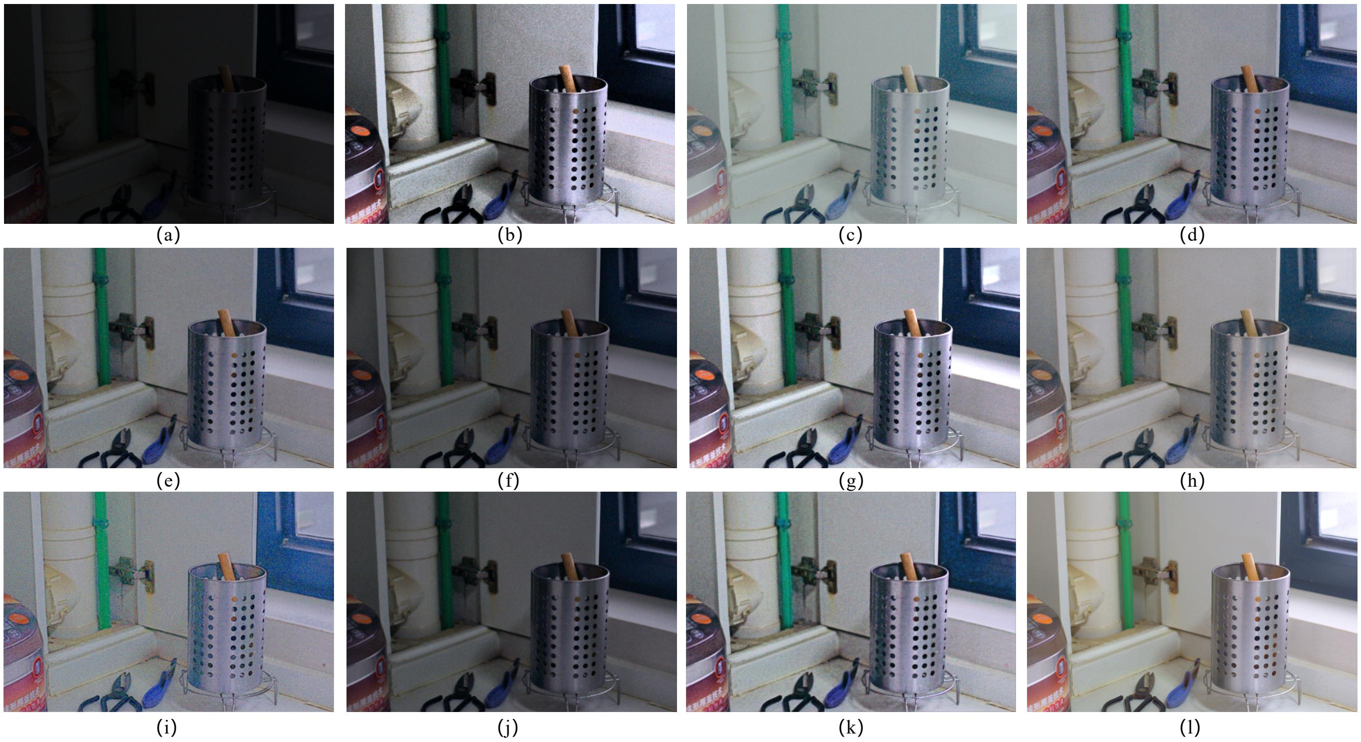

Fig. 8. The enhancement results by different methods (a) Original. (b) HE. (c) MSR (d) NPE. (e) MF. (f) SRIE. (g) LIME. (h) Gladnet. (i) Retinex-Net. (j) L_2-L_p. (k) Ours. (l) Reference.

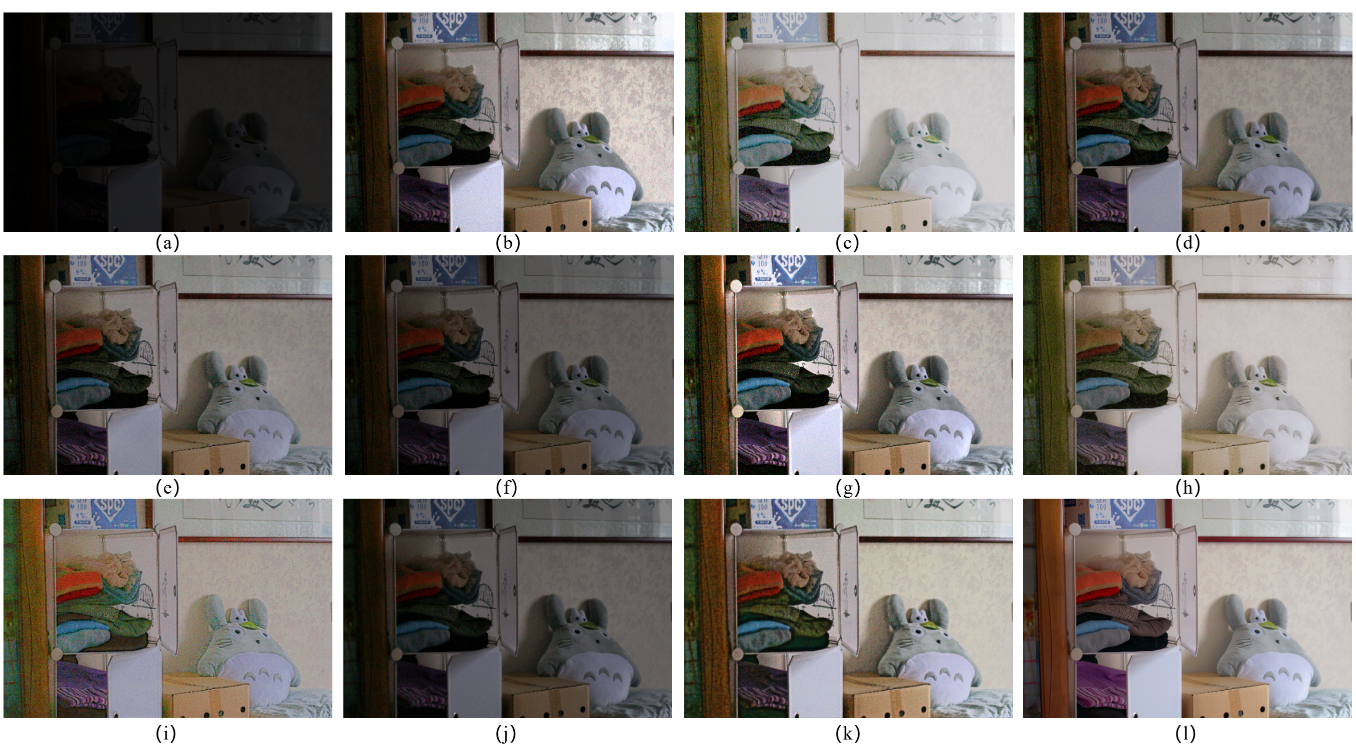

Fig. 9. The enhancement results by different methods (a) Original. (b) HE. (c) MSR (d) NPE. (e) MF. (f) SRIE. (g) LIME. (h) Gladnet. (i) Retinex-Net. (j) L_2-L_p. (k) Ours. (l) Reference.

Fig. 10. The enhancement results by different methods (a) Original. (b) HE. (c) MSR (d) NPE. (e) MF. (f) SRIE. (g) LIME. (h) Gladnet. (i) Retinex-Net. (j) L_2-L_p. (k) Ours:Epoch=120. (l) Ours:Epoch=200.

image. The training image is one of the test data of the LOL dataset and the test data is all the 15 test data of LOL dataset. In Fig.12, we use image 12-(a) only to train the network, and 12-(b) to 12-(k) are the enhancement results of image 12-(a) by different training epochs. Fig.11 displays the evaluation indexes on the test data by different training epochs. Fig.13 shows the enhancement results on some LOL test data and other low light images. Table 2 shows the indexes on the test data by 10000 training epochs. Through Fig.11 to Fig.13, we can see that our method can be applied to new environments quickly, even if we only have one image of the new environment.

It can be seen that in terms of visual effect and some indexes, the result of single image training is worse than that of multiple images training. However in our experiment, with the increase of training times, single image training does not produce artifacts. That can prove that the artifacts are not produced by the model proposed in this paper. As we train the image with single low light image, there is no need for the network to fit the histogram equalization stretch, that is an main reason why there are no artifacts in single low light training. It is perhaps difficult to fit the histogram equalization stretch when the network is trained without considering the whole image information in multiple images training. We think that if we want to avoid the artifacts in multiple images training, we have to make the network deeper or considering the whole image information, however, that will also increase time consumption.

5 Conclusion

In this paper, we propose a maximum entropy based Retinex model and a self-supervised image enhancement network. The network can be trained with low light images only and can slightly reduce the noise during enhancement. By testing on real low light images, it shows that, with short time training, the network can produce a well visual effect and has a good real-time performance. It should be noted that our method is self-supervised, so it can adapt to new environments and devices, also, the enhanced image may be different from the real data and look more like the night one in color. The future work will focus on the color restoration, noise and artifact suppression, better detail keeping and so on, we think those can be achieved through Generative Adversarial Networks or new constrains.

References

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[3] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

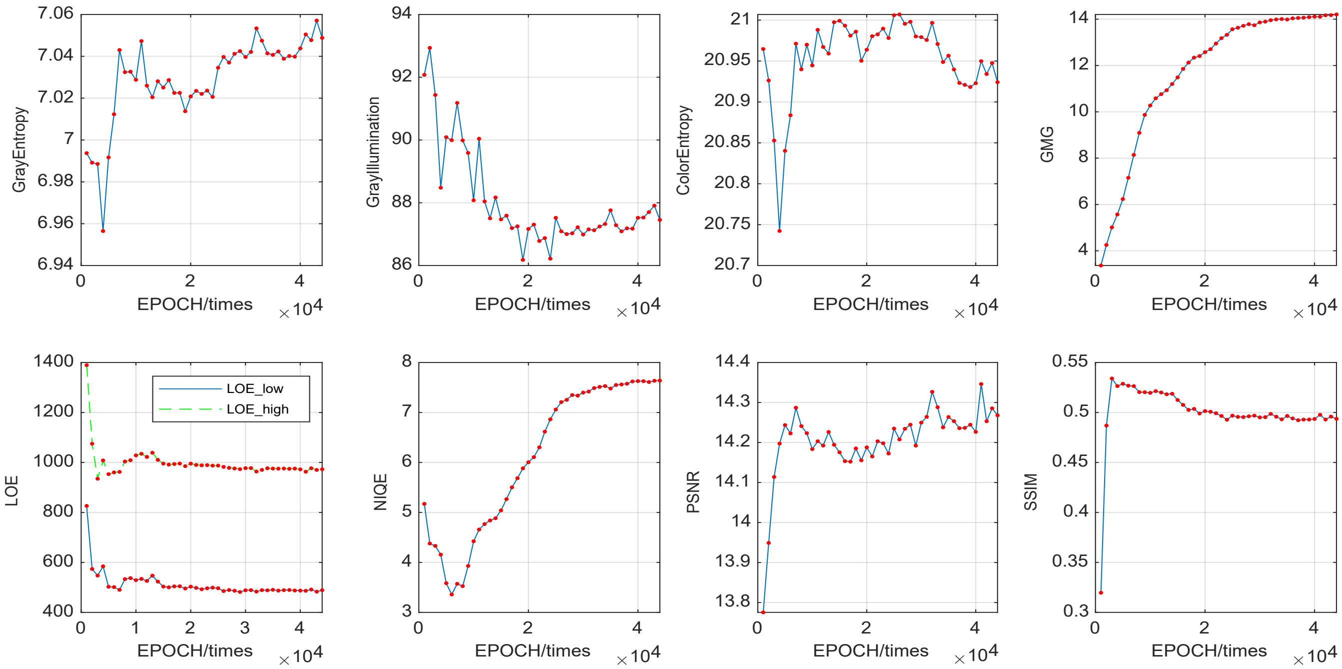

Fig. 11. Evaluation indexes on the testing data by different training epochs. The network is training with single low light image.

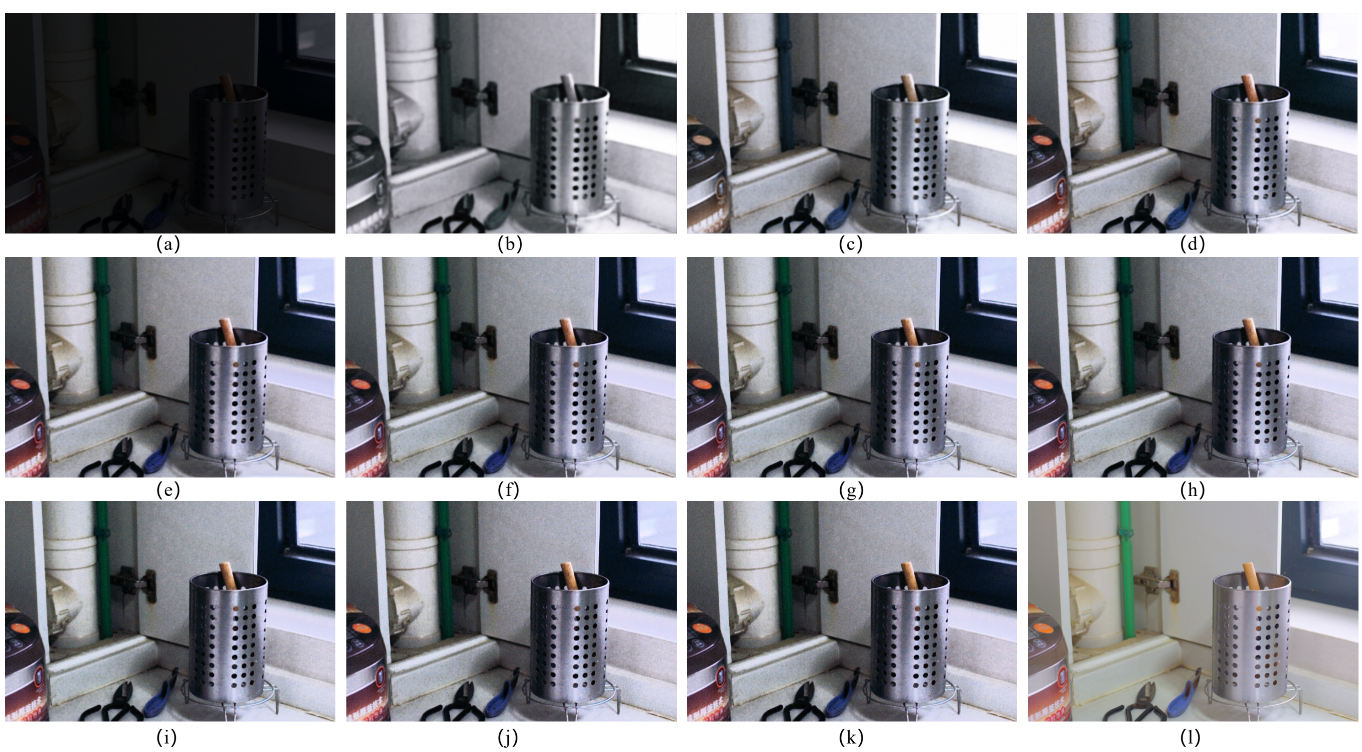

Fig. 12. The enhancement results with different training times and the network is trained with image (a) only. (a) Original. (b)Epoch=1000. (c) Epoch=2000. (d) Epoch=3000. (e) Epoch=4000. (f) Epoch=5000. (g) Epoch=6000. (h) Epoch=7000. (i) Epoch=8000. (j) Epoch=9000. (k) Epoch=10000. (l) Reference.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[7] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2015.

[8] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.

[9] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[10] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, vol. 6, pp. 22 084–22 093, 2018.

[11] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[12] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital imaging*, vol. 11, no. 4, p. 193, 1998.

[13] S. K. Naik and C. Murthy, "Hue-preserving color image enhancement without gamut problem," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1591–1598, 2003.

[14] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, 1997.

[15] Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 1, pp. 68–75, 1999.

[16] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, 2011.

[17] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[18] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu, "Fast efficient algorithm for enhancement of low lighting video," in *2011 IEEE International Conference on Multimedia and Expo*. IEEE, 2011, pp. 1–6.

[19] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[20] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 3730–3734.

[21] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE transactions on image processing*, vol. 6, no. 3, pp. 451–462, 1997.

[22] ——, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image processing*, vol. 6, no. 7, pp. 965–976, 1997.

[23] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[24] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[25] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[26] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *International Journal of computer vision*, vol. 52, no. 1, pp. 7–23, 2003.

[27] X. Fu, D. Zeng, Y. Huang, X. Ding, and X.-P. Zhang, "A variational framework for single low light image enhancement using bright channel prior," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 1085–1088.

[28] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, "Low-light image enhancement using variational optimization-based retinex model," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 2, pp. 178–184, 2017.

[29] G. Fu, L. Duan, and C. Xiao, "A hybrid l2-lp variational model for single low-light image enhancement with bright channel prior," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1925–1929.

[30] R. L. Lagendijk and J. Biemond, *Iterative identification and restoration of images*. Springer Science & Business Media, 2012, vol. 118.

[31] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[32] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1712–1722.

[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[34] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3867–3876.

[35] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3883–3891.

[36] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.

[37] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[38] C. Li, J. Guo, F. Porikli, and Y. Pang, "Lightennet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognition Letters*, vol. 104, pp. 15–22, 2018.

[39] J. Yang, X. Jiang, C. Pan, and C.-L. Liu, "Enhancement of low light level images with coupled dictionary learning," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 751–756.

[40] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *arXiv preprint arXiv:1711.02488*, 2017.

[41] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[42] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1632–1640.

[43] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[44] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[46] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[47] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2782–2790.

[48] W. Wang, C. Wei, W. Yang, and J. Liu, "Gladnet: Low-light enhancement network with global awareness," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 751–755.

Yu Zhang was born in Hebei, China. He received the B.E. degree and M.S.degree in Control Science and Engineering from Harbin Institute of Technology, China, in 2016 and 2018, respectively. He is currently studying for a Ph.D degree in Electronic science and technology in National Key laboratory of Tunable Laser Technology, Harbin Institute of Technology. His research interests include image restoration and enhancement, SLAM.

Xiaoguang Di was born in Heilongjiang,China. He received the M.S. and Ph.D degree in navigation, guidance and control from Northwestern Polytechnical University, China, in 1999 and 2004, respectively. He is currently an associate professor with the Control and Simulation Center, Harbin Institute of Technology, Where he is in charge of courses in digital image processing and computer vision. His current research interests include real-time image restoration and enhancement, 3D object detection and recognition, SLAM. Prof. Di is a member of China Simulation Federation and Chinese Society of Astronautics.

Bin Zhang was born in Gansu, China. He received the B.Sc. degree in apply physics and the M.Sc. degree in physics from China University of Petroleum, in 2010 and 2013, respectively. He is currently studying for a PhD degree in National Key Laboratory of Tunable Laser Technology, Harbin Institute of Technology. His research interests include laser imaging and laser transmission.

Chunhui Wang Chunhui Wang received the BS, MS, and PhD degrees from Harbin Institute of Technology in 1987, 1991, and 2005, respectively. He is currently a processor and Deputy Director, Institute of Optoelectronic Technology, Harbin Institute of Technology. His current major research interests are laser remote sensing, lidar, laser detection and recognition.

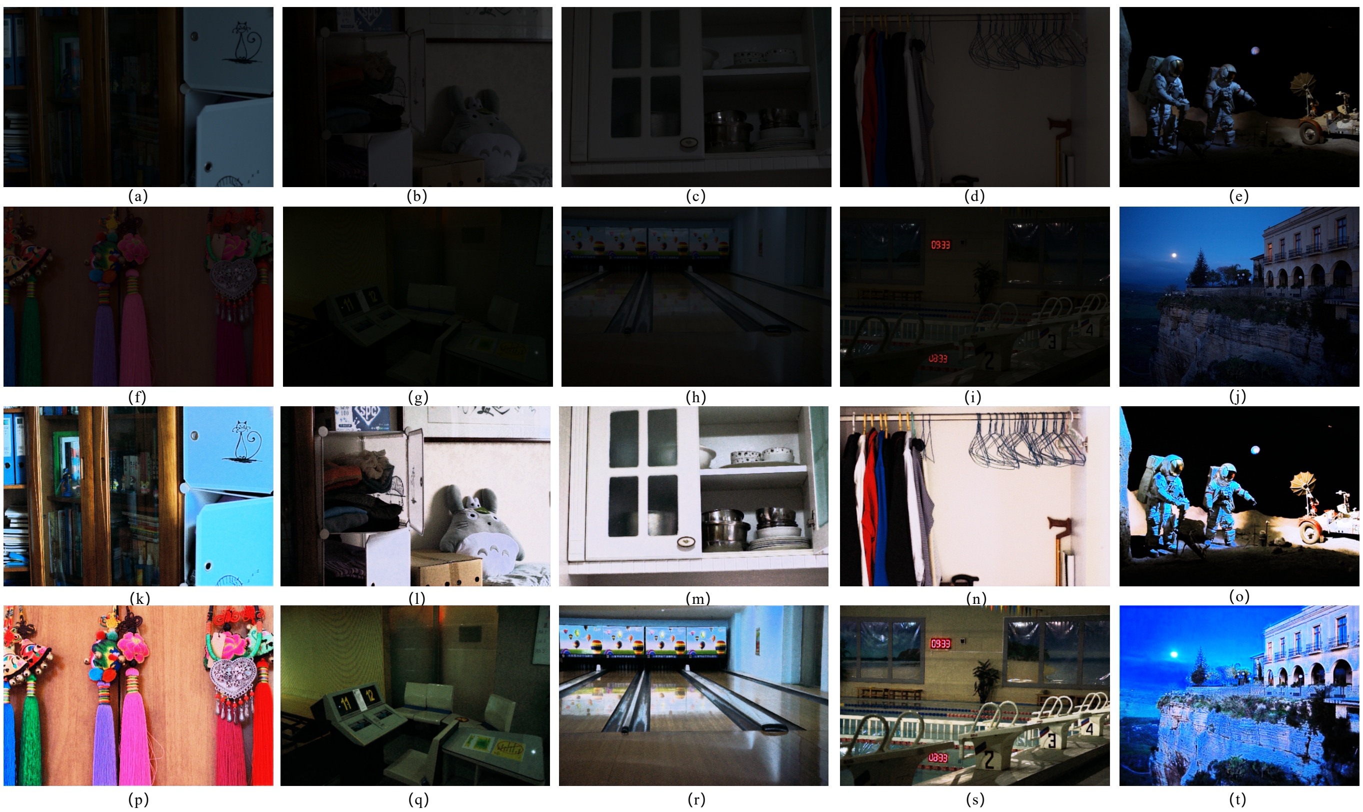

Fig. 13. The enhancement results and the network is trained by 10000 epochs with image 12-(a) only. (a)-(h) are original low light images. (i)-(p) are enhancement results.